Check for updates

RESEARCH ARTICLE

## REVISED Date of introduction and epidemiologic patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020 [version 2; peer review: 2 approved]

Mihaly Koltai [ID][1], Abdihamid Warsame[1], Farah Bashiir[2], Terri Freemantle[3], Chris Reeve[3], Chris Williams[3], Mark Jit[1], Stefan Flasche [ID][1], Nicholas G. Davies[1], CMMID COVID-19 working group, Ahmed Aweis[3], Mohamed Ahmed[3], Abdirisak Dalmar[2], Francesco Checchi[1]

[1]London School of Hygiene & Tropical Medicine, London, UK
[2]Somali Disaster Resilience Institute, Mogadishu, Somalia
[3]Satellite Applications Catapult, Didcot, UK

## Abstract

**Background:** In countries with weak surveillance systems, confirmed coronavirus disease 2019 (COVID-19) deaths are likely to underestimate the pandemic's death toll. Many countries also have incomplete vital registration systems, hampering excess mortality estimation. Here, we fitted a dynamic transmission model to satellite imagery data of cemeteries in Mogadishu, Somalia during 2020 to estimate the date of introduction and other epidemiologic parameters of the early spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in this low-income, crisis-affected setting.

**Methods:** We performed Markov chain Monte Carlo (MCMC) fitting with an age-structured compartmental COVID-19 model to provide median estimates and credible intervals for the date of introduction, the basic reproduction number ($R_0$) and the effect of non-pharmaceutical interventions (NPIs) up to August 2020.

**Results:** Under the assumption that excess deaths in Mogadishu March-August 2020 were attributable to SARS-CoV-2 infections, we arrived at median estimates of November-December 2019 for the date of introduction and low $R_0$ estimates (1.4-1.7) reflecting the slow and early rise and long plateau of excess deaths. The date of introduction, the amount of external seeding, the infection fatality rate (IFR) and the effectiveness of NPIs are correlated parameters and not separately

## Open Peer Review

**Approval Status** ✓ ✓

|  | 1 | 2 |
|---|---|---|
| **version 2** (revision) 18 Feb 2022 |  | ✓ view |
|  |  | ↑ |
| **version 1** 06 Oct 2021 | ✓ view | ? view |

1. **Zindoga Mukandavire**, Emirates Aviation University, Dubai, United Arab Emirates

2. **Bhaswar Ghosh**, International Institute of Information Technology, Hyderabad, India

Any reports and responses or comments on the article can be found at the end of the article.

identifiable in a narrow range from deaths data. Nevertheless, to obtain introduction dates no earlier than November 2019 a higher population-wide IFR ($\geq$0.7%) had to be assumed than obtained by applying age-specific IFRs from high-income countries to Somalia's age structure.

**Conclusions:** Model fitting of excess mortality data across a range of plausible values of the IFR and the amount of external seeding suggests an early SARS-CoV-2 introduction event may have occurred in Somalia in November-December 2019. Transmissibility in the first epidemic wave was estimated to be lower than in European settings. Alternatively, there was another, unidentified source of sustained excess mortality in Mogadishu from March to August 2020.

### Keywords
COVID-19, SARS-CoV-2, transmission model, excess mortality, underascertainment, Somalia, COVID-19 in low-income countries
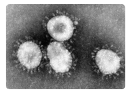
This article is included in the Coronavirus (COVID-19) collection.

**Corresponding authors:** Mihaly Koltai (mihaly.koltai@lshtm.ac.uk), Francesco Checchi (francesco.checchi@lshtm.ac.uk)

**Author roles: Koltai M**: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Warsame A**: Data Curation, Resources, Writing – Review & Editing; **Bashiir F**: Data Curation, Resources; **Freemantle T**: Data Curation, Resources; **Reeve C**: Data Curation, Resources; **Williams C**: Data Curation, Resources; **Jit M**: Conceptualization, Writing – Review & Editing; **Flasche S**: Conceptualization, Writing – Review & Editing; **Davies NG**: Resources; **Aweis A**: Data Curation, Resources; **Ahmed M**: Data Curation, Resources; **Dalmar A**: Data Curation, Resources; **Checchi F**: Conceptualization, Funding Acquisition, Project Administration, Resources, Writing – Review & Editing

> **REVISED**  **Amendments from Version 1**
>
> In response to reviewers a few minor additions were added to the article.
>
> These are:
>
> - an explanation of why R0 is lower for the fitting with the lowest assumed IFR value (0.14%)
>
> - expanding the legend of Figure 4, explaining that the coloured lines correspond to separate fits with different assumes IFR values, with the solid lines corresponding to the median of the posterior distributions and the shaded regions around their CI95s
>
> - adding a column name for the first column of Table 1
>
> **Any further responses from the reviewers can be found at the end of the article**

## Introduction

By September 2021, more than 4.7 million people were confirmed to have died from the coronavirus disease 2019 (COVID-19) pandemic caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). While the cumulative rate of confirmed deaths has exceeded 1-2 per 1000 persons in several European countries, the United States and some of Latin America, it has remained one or even two orders of magnitude lower in most of Africa.

While some of this difference can be potentially explained by a lower infection fatality ratio (IFR) for the entire population due to a lower median age[1–3], evidence suggests that at least critically ill COVID-19 African patients experience higher, not lower mortality than elsewhere[4], as plausibly expected due to weaker health infrastructure[5]. News reports, studies using seroprevalence[6,7], polymerase chain reaction (PCR) testing in morgues[8], as well as indirect data sources such as obituaries on social media[9] point to substantial under-ascertainment of cases and deaths in low-income countries, potentially ten-fold (suggested by excess mortality data from Egypt[10]) and in some crisis-ridden regions perhaps even hundred-fold[11]. While in high income countries confirmed COVID-19 deaths are approximately in line with excess death statistics[10], in many African countries there are no reliable mortality statistics, precluding the use of excess death data to infer the true scale of the pandemic.

Additionally, while the first COVID-19 cases in sub-Saharan African countries were identified in late February[12] (16th of March in Somalia), there is considerable uncertainty about the true date of introduction, often estimated to be in January 2020 for Western Europe[13], or as early as December 2019 according to retrospective PCR on routine patient samples[14]. For these reasons, alternative data sources such as obituaries[9] and satellite imagery[15] of cemeteries have been leveraged to estimate the true scale of COVID-19 mortality and its early spread in low- and middle-income countries in Africa and elsewhere.

Burial data derived from cemetery records or from satellite imagery have been used by multiple studies of the COVID-19 pandemic in low-income countries. Analysis of burial records from Ethiopia's capital, Addis Ababa, showed significant excess mortality in the third quarter of 2020[16]. Analysis of mortality records from Jakarta, Indonesia[17] showed a rise in excess mortality starting in January-February 2020, more than two months before reported COVID-19 deaths, and an approximately 60% increase in deaths in the first 10 months of 2020 compared to the preceding 5 years. In a study[15] of the Aden governorate in Yemen, very high-resolution satellite images of all active cemeteries and Civil Registry office records of deaths were compared, providing validation for the use of satellite imagery. This study found substantial excess mortality in the period April–July 2020, unreported in COVID-19 statistics, with the satellite and registry data showing similar trends, but estimates from satellite imagery approximately 40% lower at the time of peak mortality. The interpolation method from this study to impute graves in the absence of sufficient resolution was also used in the current study.

In the current study we used a dynamic transmission model to analyse a time series of excess deaths in Mogadishu (Somalia) inferred from satellite images of the six main cemeteries in the city[18]. Our aim was to estimate the probable date of introduction of SARS-CoV-2, as well as the basic reproduction number ($R_0$) and the effect of non-pharmaceutical interventions. While we could not verify estimates from satellite imagery with alternative data sources due to the unavailability of death registry data, similar patterns of excess mortality were inferred in other low-income countries based on data from cemeteries[15–17], and the similar study on Yemen[15] suggests satellite imagery captures well the trend (if less the absolute numbers) of excess mortality when compared with death notifications to civil registries. While the absolute number of deaths might not be very accurately estimated by satellite imagery it nevertheless remains an important, and in the case of Somalia, to date, the only source to reconstruct the trajectory of the mortality trends of the initial stages of the COVID-19 pandemic.

## Methods
### Ethical approval
This is a modelling study on previously acquired data. There was no primary data acquisition within the framework of the study. For the original underlying data ethical approval was received from the ethics review committee of the London School of Hygiene & Tropical Medicine (REF: 22458) as well as the ethics review committee of the Somali Disaster Resilience Institute (REF: RB-0123).

### Data sources and statistical analysis
All analysis was performed using the free software R (version 3.6.3) used in Rstudio (version 1.3.1093). The analysis code has been deposited on Zenodo[19].

Details of the method for inferring excess mortality from satellite images are described in an accompanying article[18]. Briefly, cemeteries in the Banadir administrative region, which contains Mogadishu, were identified via open-source location data and satellite imagery (OpenStreetMap, Google Earth, GoogleMaps) by a combination of automatic image recognition

and manual annotation, in addition to key informant interviews and field visits to identified cemeteries. We identified and analysed six cemeteries (Barakaat 1 and 2, Calamada, Iskool Bolisii, Kahda, Moallim Nuur). We excluded from the analysis five smaller private and family-owned cemeteries estimated to account for less than 20% of all burials because of lack of images and vegetation cover. One cemetery (Calamada) included in the analysis falls outside the limits of Banadir, but largely caters to Mogadishu residents and was therefore included.

In total, 68 archive satellite images from the period February 2017-September 2020 were selected on the criteria that they were cloud-free, of high radiometric quality and with a spatial resolution of 30–40 cm per pixel, were analysed through manual and semi-automated image processing to extract surface area and number of graves. An exhaustive grave count by either of these two methods was possible for 40 out of 68 satellite (58.8%) images. For the remaining images, the number of graves was extrapolated from visible areas or imputed through a generalised additive mixed model of the association between graves and surface area. Results for each image were then interpolated and summed across all cemeteries to yield a single time series of burials for the city.

To compute the baseline (pre-pandemic) crude death rate (CDR), population denominators for Mogadishu (Banadir region) were estimated using the WorldPop project's database[20], using either the 2015 or 2019 estimates, while also adjusting for in- and out-displacement to/from the city[21]. The two alternative base estimates correspond to a 'high' and a 'low' scenario with nearly identical trends (*Extended data,* SI Figure 1[22]) and marginally different levels (0.04–0.05 deaths/10.000 person-days) of baseline (i.e. pre-pandemic) CDR.

This level is significantly lower than previous CDR estimates for Somalia[23] that were between 0.2–0.6/10.000 person-days. Assuming that the level of under-estimation remains constant, we can scale the crude death rate estimated from our time series up to previous estimates to 0.1–0.4 deaths per 10.000 person-days, using the lower half of the estimates from [23]. In terms of modelling transmission dynamics, such scaling of deaths merely shifts the IFR while leaving other parameter estimates unchanged, and hence provides little additional information. We therefore used the observed time series of burials directly for model fitting, without scaling. To isolate excess mortality (which we assumed to be entirely attributable to SARS-CoV-2 infection: see Discussion), we extrapolated pre-2020 burial rates into the pandemic period and subtracted this baseline from the total (*Extended data,* SI Figure 1[22]). Specifically, to calculate excess burials (mortality) we took the daily number of burials in the dataset and subtracted the mean level of daily burials in the four months period 01/07/2019-01/11/2019 (9.3 burials/day). We chose this limited pre-pandemic period as a basis of comparison since burial rates in the preceding period had been likely affected by the drought-triggered crisis 2017-2018[23]. The model output of incident deaths was fitted to this baseline-subtracted number of burials per day.

## Transmission model

We used CovidM, an age-stratified dynamic transmission model initially developed to model the spread of COVID-19 and the effect of non-pharmaceutical interventions in the UK[24,25]. The model has a susceptible-exposed-infectious-recovered structure with individuals stratified into 5-year age bands. When susceptible individuals are infected they move into an exposed (incubating) compartment (E), becoming either infectious with symptoms ($I_c$) following a pre-symptomatic phase ($I_p$), or remaining asymptomatic ($I_s$) with a lower level of infectiousness (set to 50% as in previous studies[24,26]). We used existing age-dependent estimates[24] for the proportion of individuals who are symptomatically infected (clinical fraction), as well as for the susceptibility to infection (*Extended data*, SI Figure 5[22]). Both of these estimates are age-dependent, with the clinical fraction 29% in the 0–9y age group and 69% above 70 years, and susceptibility among individuals aged 0–19y half of that among adults. Deaths occur in the model with a gamma-distributed delay[27,28] following the transition from exposed (E) to pre-symptomatic ($I_p$) state. To account for the lack of intensive care unit (ICU) beds (the available estimate is around twenty ICU beds in the entire Mogadishu area[18]), we set the mean delay to 15 days, one week less than the estimate used for high-income countries[26], to reflect the fact that severe cases were unlikely to receive adequate treatment. Other parameters of disease progression were fixed to consensus estimates in the literature (*Extended data*, SI Table 1[22]). The model was parameterised with the demographic structure of Somalia[20] and the total population of Mogadishu (2.2 million as of mid-2020). Since there is no empirical contact matrix available for Somalia, we used the projected contact matrix[29] for its neighbour Ethiopia.

## Estimates on infection fatality ratio

We used existing age-specific IFR estimates[26] demonstrating a log-linear relationship between age and the IFR. These estimates are from high-income countries and the IFR is likely to be different in Somalia. To account for this difference and also because the IFR cannot be identified from death data only, we fitted the data with multiple IFR estimates. The estimates were shifted upward to reflect the effect of a weak public health infrastructure. To do so, we took the logit of the original IFR at each age group and increased its value (*Extended data,* SI Figure 6[22]), raising the mean IFR for those 75 or older from the original 11.6% to the range of 26–70%. A possible upward shift of IFR values by age groups is supported by recent findings of substantially higher in-hospital mortality in several African countries[4] and by mortality estimates from auxiliary sources and modelling[30]. The population-average IFR (calculated for a randomly chosen infected person) for Somalia would be 0.15% with direct transferring the original estimates and 0.35%, 0.77%, 1.1% and 1.6%, respectively, with the adjusted values (logit increased by 1, 2, 2.5 and 3, respectively; *Extended data,* SI Figure 6[22]).

## Input and fitting parameters

We estimated three epidemiological parameters of the model: the date of introduction into Mogadishu, the basic reproduction number ($R_0$) and the effectiveness of the NPIs converting the

nominal stringency of non-pharmaceutical interventions (NPIs) into the actual reduction of transmissibility. We held two other parameters, the IFR and the size of the initial seeding event (number of incubating infected individuals (E) that entered the region), fixed and performed the fitting for a range of values. The IFR cannot be determined from death data only, whereas the seed size and the date of introduction are inversely related parameters and not separately identifiable. Therefore, we fitted only the three parameters mentioned, for a range of different IFR levels and seed sizes between 20 and 200 (we tested even larger seed sizes; *Extended data,* SI Figure 11[22]). For simplicity we placed the seeding event onto a single day; in our deterministic modelling framework a more gradual introduction does not have a significantly different effect. Initial importations were assumed to be 30 to 70 year-old adults.

The effect of NPIs was accounted for by using the Oxford COVID-19 Government Response Tracker (OxCGRT)[31], using the *StringencyIndex* variable for the strength of the NPIs. This estimate is based on mitigation policies announced, which we expect not to be entirely (or even significantly) effective. Since we have no independent data source (such as mobility data) on the actual effect of NPIs in Somalia, we introduced a fitting parameter (*NPI_scale*) representing the effectiveness of NPIs that converts the nominal stringency into actual contact reductions. We distinguished three periods in the NPIs (*Extended data,* SI Figure 3[22]). In the first, the value of *StringencyIndex* increased abruptly (in three days) from 0 to 41% of its maximum on the 18th of March and stayed above 50% until the 30th of June. From the 1st of July to the 29th of August a number of relaxations followed (second period). In the third period, from the 30th of August the *StringencyIndex* started to increase again and did not decrease until the end of our fitting period. The third period is outside the time window of our model fitting. To minimise over-parameterisation, instead of using the full time series of *StringencyIndex* that would introduce significant additional complexity to model dynamics we only took the mean value of *StringencyIndex* in these three periods (0.59, 0.26 and 0.41; *Extended data,* SI Figure 3[22]), and implemented the effect of NPIs by reducing transmission coefficients for all age groups by the product of the mean value *StringencyIndex* per period and the NPI effectiveness (*NPI_scale*). For example, if *NPI_scale*=0.5, then the actual reduction in transmission was 29.5% (since it is the product *StringencyIndex*NPI_scale*= 0.59*0.5) in the first period when stringency was 0.59, and 13% (*StringencyIndex*NPI_scale*=0.26*0.5) in the second period.

### Time window of fitting
All fits presented in the main text were done with data within the time window 02/03/2020 to 24/08/2020, excluding the first smaller spike of deaths in January-February, as well as the late spike in September. We removed the January–February spike in excess burials to avoid any confounding from the two continuous weeks of reported cholera deaths in Banadir[31] that coincided with this period (see Discussion). Moreover, this early spike of deaths is in general inconsistent with a gradually rising epidemic curve from the beginning of March. Including the

deaths in January–February leads to even earlier estimates for the date of introduction, but poorer fits (*Extended data,* SI Figure 10[22]), as the epidemiological model cannot capture this early non-monotonic dynamics.

### Fitting procedure
To estimate the unknown parameters, we fitted the CovidM model to the time series of excess deaths using a Monte Carlo Markov chain (MCMC) algorithm, minimising the log-likelihood of incident deaths (assumed to be Poisson-distributed). We introduced informative prior assumptions for the date of introduction (normal distribution with mean: 01/03/2020, standard deviation: 20 days) and $R_0$ (truncated normal distribution with mean=3, standard deviation=1, bounded at 1 and 5), and an uninformative uniform distribution for the NPI scaling factor (U(0,1)). We used a differential evolution MCMC algorithm with 10 chains, with a burn-in of at least 500 iterations followed by at least 2000 samples.

### Results
Satellite imagery of the six main cemeteries in Mogadishu showed a first spike in the number of burials in late January and February, followed by a more sustained rise from early March (Figure 1). The weekly number of excess burials rose to approximately 60 in April and to a peak of 85 in mid-June, falling back to values near zero only in August[32].

### Estimates for date of introduction and initial spread ($R_0$)
The slow rise in deaths from March to mid-June and the long plateau lasting until late July results in $R_0$ estimates substantially lower than those for Wuhan[33] and European countries[25] in the initial phase of the pandemic. Fitting our data with IFR values between 0.15% and 1.6% and seed sizes from 20 to 200 resulted in median $R_0$ estimates of 1.4-1.5 (Figure 2–Figure 3, Table 1). Fit quality as expressed by DIC (deviance information criterion) values are similar for population-wide IFRs between 0.36% and 1.6% (Figure 3), but much poorer when using the high-income country-specific base assumption of 0.15%.

Using population-wide IFR values below 0.5% leads to implausibly early date of introduction estimates before November 2019 (Figure 3). The lowest IFR value also leads to a higher $R_0$ estimate, as a larger wave of infections is required to fit the number of deaths in the data, however, this fit is unable to account for the dynamics of deaths (Figure 4). With progressively higher assumed population-wide IFRs, the median date of introduction estimate gradually moves to mid- and late November 2019, or early December when assuming a large (200 introductions) seeding event. Even larger seeding events of 500 or 1000 introductions shift the median date of introduction estimates to mid-December for the highest IFR values (*Extended data,* SI Figure 11[22]). In other words, our excess mortality time series can be fitted either with an early introduction date or a very large amount of external seeding at a later date. However, even in the case of up to 1000 external introductions, the date of introduction is still mid-December 2019, earlier than most previous estimates.
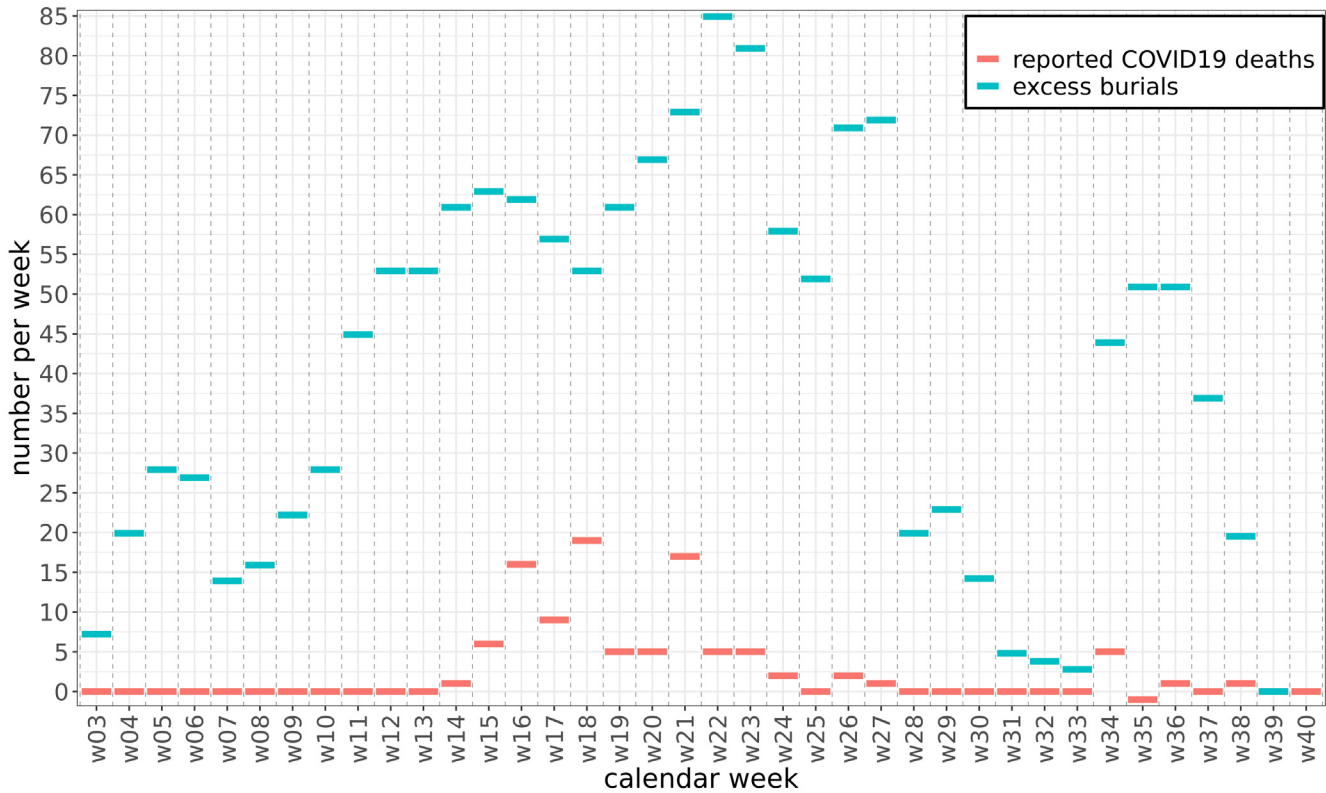
**Figure 1. Excess burials and reported coronavirus disease 2019 (COVID-19) deaths.** Weekly burials above the pre-pandemic baseline (excess burials) in Mogadishu compared to reported COVID-19 deaths in Somalia.

### Estimates for the effect of NPIs

Estimates for the effectiveness of NPIs show positive correlation with the population-average IFR and (weaker) negative correlation with the seed size (Figure 2). For the first NPI period, when NPIs were the most stringent (19/March to 30/June), we obtained median estimates for the NPI-caused transmissibility reductions between 25 and 30% for the IFR values (Figure 3) 0.77% and above. These values are lower than NPI-induced reductions of contact rates in high income countries[34], however they do have an effect on the growth of cases for the fits with IFR values of 0.77% and above, as they break the exponential growth of cases, resulting in the long plateau of deaths observed in our burial data (Figure 4–Figure 5). This breakpoint in the dynamics of infections due to the stronger effect of NPIs for the fits with IFR=0.77% and above also shifts the date of introduction to mid- to late November 2019 (Figure 2 and Figure 4), whereas the fits for lower IFRs yield lower estimates for NPI effectiveness but implausibly early introduction dates (October 2019).

### Uncertainty in estimates due to parameter correlations

There are strong positive correlations between the three fitting parameters (*Extended data,* SI Figure 8[22]), nevertheless 50% and 95% credible intervals for all three are relatively narrow (Figure 2). A higher assumed IFR value and a larger seed size both shift the date of introduction to later dates, and larger

assumed seed sizes also lower estimates for the basic reproduction number (Figure 2–Figure 3). The quality of fits in terms of DIC values are essentially the same for the IFR input values of 0.35% and above (Figure 3). Seed sizes larger than 200 do not improve the fit quality further (*Extended data,* SI Figure 11[22]) while being less likely themselves.

### Estimates using only the initial phase of excess deaths

The early estimates of the date of introduction are not only due to the early appearance of excess deaths in our data, but also due the low $R_0$ estimates, which are in turn affected by the long plateau of deaths over the summer of 2020 after a first exponential phase. Estimates of the number of burials from satellite imagery in Yemen[15] indicated differential underestimation of deaths (compared to death notifications to the civil registry) at the peak of excess mortality, although still correctly reflecting the general trend of excess mortality. Therefore, it is possible that the real peak of deaths was higher and the excess mortality curve had a sharper exponential rise, but the number of burials underestimates excess deaths during the peak period of the pandemic, perhaps because of out-migration from the city (as observed in India[35]) during the pandemic or opening of new burial sites not included in our satellite data. While field visits and interviews[18] did not identify new burial sites and we therefore cannot ascertain the veracity of this hypothesis, we approximated it by re-fitting the model to the
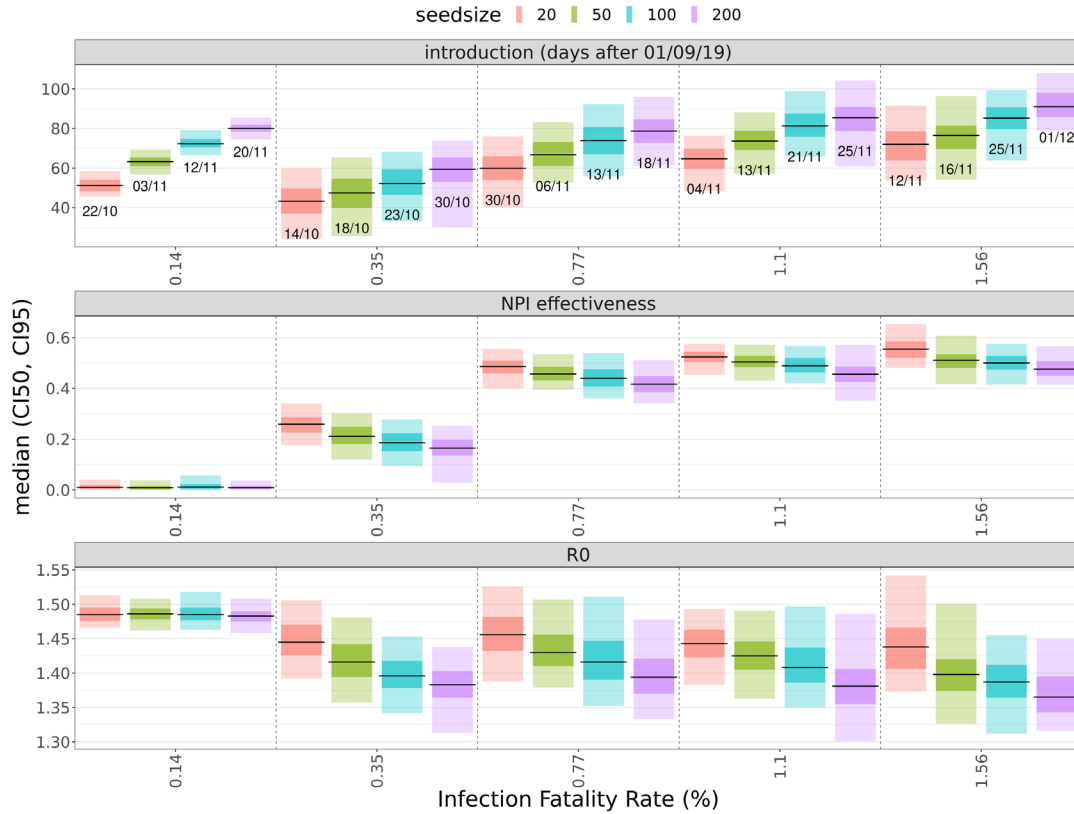
**Figure 2. Estimates of fitting parameters.** Median values and credible intervals for the fitting parameters (introduction date, NPI effectiveness [NPI_scale], $R_0$) and quality of fits at different assumed values of the infection fatality rate (x-axis) and seed size (colors). In the top panel, labels below the lines show median estimates of the date of introduction (dd/mm, all dates are in 2019). Shaded areas around the median (black) are 50% (darker) and 95% credible intervals.
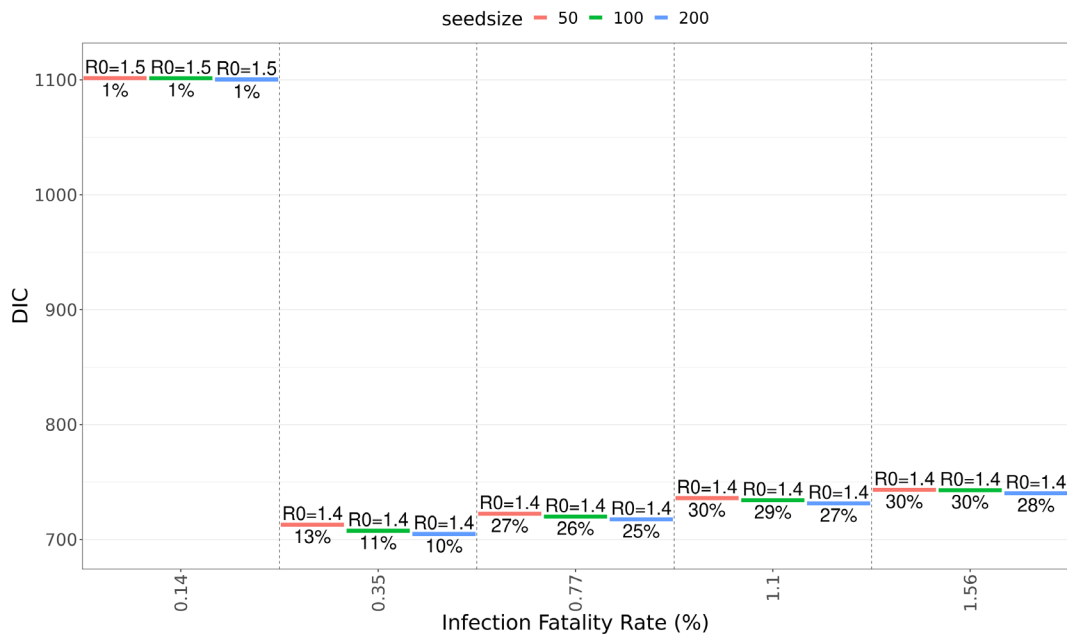


**Figure 3. Quality of fits for different infection fatality ratios and seed sizes.** Goodness of fit as measured by DIC (deviance information criterion) at different values of seed size and population-wide IFR (infection fatality rate). Labels above the coloured lines show median estimates for $R_0$, labels below show the NPI-induced reduction in transmissibility (during the first NPI period).

**Table 1. Parameter estimates.** Mean values and 95% credible intervals of the three fitting parameters, for different values of the infection fatality rate (IFR) and the seed size. NPI=non-pharmaceutical interventions.

| Parameter description | IFR=0.77% | IFR=1.1% | IFR=1.6% | seed size |
|---|---|---|---|---|
| date of introduction (dd/mm, all dates 2019) | 30/10 (11/10, 16/11) | 04/11 (19/10,16/11) | 12/11 (24/10,01/12) | 20 |
| | 06/11 (23/10, 23/11) | 13/11 (28/10,28/11) | 16/11 (25/10,06/12) | 50 |
| | 13/11 (26/10, 02/12) | 21/11 (05/11,08/12) | 25/11 (03/11,09/12) | 100 |
| | 18/11 (01/11, 05/12) | 25/11 (31/10,14/12) | 01/12 (19/11,18/12) | 200 |
| $R_0$ | 1.46 (1.39,1.53) | 1.44 (1.38,1.49) | 1.44 (1.37,1.54) | 20 |
| | 1.43 (1.38,1.51) | 1.43 (1.36,1.49) | 1.4 (1.33,1.5) | 50 |
| | 1.42 (1.35,1.51) | 1.41 (1.35,1.5) | 1.39 (1.31,1.46) | 100 |
| | 1.39 (1.33,1.48) | 1.38 (1.3,1.49) | 1.36 (1.32,1.45) | 200 |
| NPI_scale (NPI effectiveness) | 0.49 (0.4,0.56) | 0.52 (0.45,0.57) | 0.56 (0.48,0.65) | 20 |
| | 0.46 (0.4,0.54) | 0.5 (0.43,0.57) | 0.51 (0.42,0.61) | 50 |
| | 0.44 (0.36,0.54) | 0.49 (0.42,0.57) | 0.5 (0.42,0.57) | 100 |
| | 0.42 (0.34,0.51) | 0.46 (0.35,0.57) | 0.48 (0.42,0.57) | 200 |

pre-peak period up to 13 April only. This led to a shift in the date of introduction estimates to the period of mid-November to mid-December for seed sizes of 20 to 100 (*Extended data,* SI Figure 9[22]), with median $R_0$ estimates rising to 1.4–1.7. Date of introduction estimates in this case are no earlier than mid-November even in the case of a seed size of 20.

### Deaths from political violence in 2020

Since Somalia is heavily affected by political violence, including armed conflicts and terror attacks[36], we also investigated if a concurrent rise in violent deaths could explain the sustained rise in burials observed March-August 2020. We analysed the number of fatalities due to political violence documented for Somalia in the Armed Conflict Location & Event Data Project (ACLED) database for the years 2018, 2019 and 2020. Compared to 2018-2019, for the year of 2020 we found no increase, but rather a reduction, in the number of fatalities due to political violence in the Banadir region (*Extended data,* SI Figure 4[22]). While there was one major terrorist attack claiming 85 lives on the 28th December 2019, this was followed by a long period of deaths below the level of the previous year, followed by two major incidents (20 and 26 deaths, respectively) in August 2020. In the period from February to July, when the daily burial rate doubled from its baseline, there was no increase of fatalities due to political violence, therefore we find no evidence supporting the hypothesis that the rise in burials could be due to this exogenous factor.

### Discussion

Fitting excess mortality in Mogadishu from the beginning of March by a validated SARS-CoV-2 age-structured compartmental model[24,26] we arrived at date of introduction estimates from mid-November to early December 2019, at least one

month earlier than previous estimates[37]. The estimate shifts to mid-December only by assuming one thousand or more external introductions (*Extended data,* SI Figure 11[22]). Additionally, our estimates of the basic reproduction number between 1.4 and 1.7 are also lower than previous estimates[38]. These two findings are not only due to the early appearance of excess mortality, but also the slow rise of deaths and their sustained plateau from April to July, leading to low $R_0$ estimates and a consequent dating-back of the introduction date to early time points. While the epidemiological model can fit the deaths data relatively well (Figure 4–Figure 5), the introduction dates of mid-November to early December 2019 are surprising based on the current understanding of the early phase of the COVID-19 pandemic, usually dating the introduction of the pathogen to Africa to January 2020[12,39].

This study has several limitations. Our model fitting of excess deaths is predicated on the strong assumption that the unexplained rise in burials from March 2020 was due to deaths caused by SARS-CoV-2 infection. We investigated a number of alternative hypotheses other than COVID-19 that could explain the observed excess mortality.

There was an ongoing cholera epidemic in Somalia following floods in 2017[31], resulting in 19 confirmed deaths in Banadir (mostly children) from January to October 2020. In the four weeks of 20 January to 06 February, approximately coinciding with the first transient increase of excess burials in our dataset, there were four confirmed cholera deaths in the whole of Somalia, after a preceding period of no reported deaths. In the period from 16 February to 12 April there were eight further cholera deaths reported Somalia-wide, and another 12 deaths in June-July. These numbers are much lower than
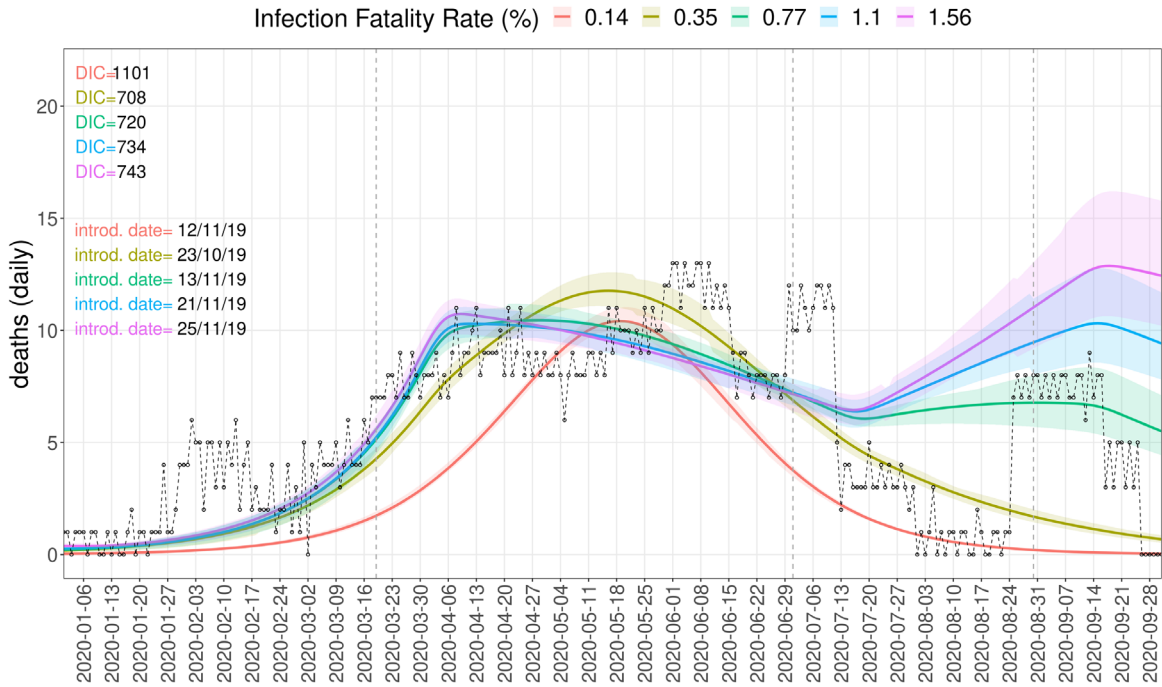
**Figure 4. Burial data and simulated deaths.** Dynamics generated by sampling the posterior distributions of fitting parameters, at a seed size of 100 and four IFR values from 0.14% to 1.56%. Different colours correspond to separate fits with different assumed IFR values, with the solid lines corresponding to the median of the posterior distribution and the shaded regions the 95% confidence intervals. The dashed black line and circles show the daily number of excess burials. The period from 2 March to 24 August was used for fitting.
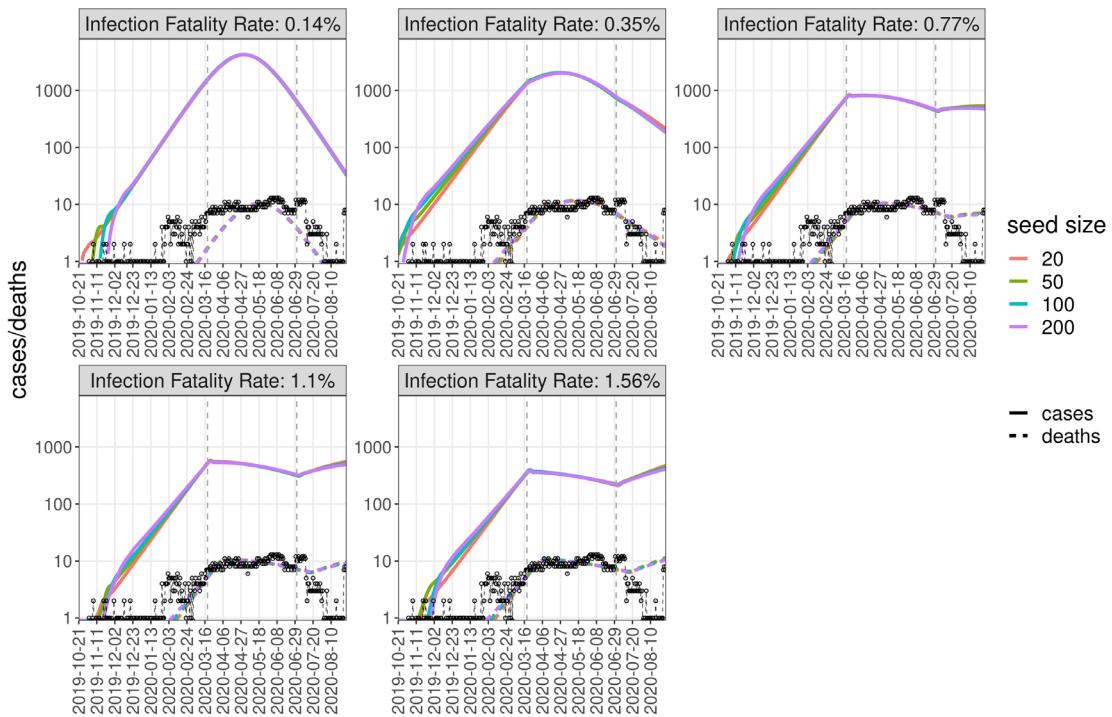


**Figure 5. Simulated dynamics of cases and deaths.** Simulated dynamics of cases (solid lines) and deaths (dashed, coloured) for different seed sizes (colours) and IFR values (subplots) using the mean values of fitted parameters, compared to daily number of excess burials (black dashed line and circles).

the observed total increase of burials between February and October 2020: approximately 1500 excess burials were directly identified from satellite imagery and we estimated total excess deaths in Banadir to be between four and twelve thousand. While some underestimation of cholera-related deaths is possible, due to its well-identifiable pathology we consider it unlikely that a major cholera outbreak was almost entirely missed and could explain a substantial proportion of the excess mortality.

There are two, non-exclusive ways to interpret these findings. On the one hand, given uncertainties about the earliest phase of the pandemic in Wuhan[40], it is possible that SARS-CoV-2 was imported to Mogadishu at an earlier date than most consensus estimates. Mogadishu, the only international airport in Somalia, received flights with over thirty thousand seats in total per month in the period before the COVID-19 pandemic[39]. The country has connecting flights with multiple countries (UAE, Turkey, Kenya, Ethiopia, Qatar) that have several daily flights with China and in two cases (Turkey, UAE) with Wuhan. Moreover, news articles reported at least 34 Somali students in Wuhan, with the entire Somali diaspora likely to be larger, as trade and general economic relations between China and Somalia have been expanding in the last two decades, resulting in a growing Somali (and other African) diaspora in China[41]. Larger neighbouring countries Kenya and Ethiopia have far more extensive trade[42] and travel flows with China, making indirect importation to Somalia possible, but implying that earlier introduction dates could have happened for those countries as well. SARS-CoV-2 positive routine samples from mid-December 2019 were also found in Italy[14], France[43] and the United States[44], suggesting the pathogen was circulating in small numbers by the end of 2019 outside of China, but resulting in excess mortality only from Spring 2020. Phylogenetic analysis[45] suggests that a progenitor of the SARS-CoV-2 variant first identified in Wuhan might have been spreading outside of China before the known beginning of the city's outbreak.

In the compartmental-deterministic framework we used, superspreading events can only be incorporated as static model inputs (i.e., an injection of cases into the model), although the large seed sizes used as input parameters can be interpreted as proxies for early superspreading events that followed smaller seeding events. In Somali society, superspreading events such as large funerals or marriages may be more likely than in Europe, such that the importation of even a few seed cases in late 2019 might have resulted in extensive early propagation.

If some of the excess mortality observed in our dataset was due to causes other than COVID-19, a more sharply rising epidemic curve might be hidden within the curve of all excess deaths, which, if disentangled, would lead to a higher $R_0$ estimate and therefore a later date of introduction. We tried to approximate this potential effect by an alternative fit of only the first, exponential phase of the excess mortality curve, but not the ensuing long plateau which might have been due to the indirect effects of the pandemic. This alternative fit resulted in arguably more plausible estimates of the date of introduction, up to mid-December 2019. Conversely, if a very early introduction

did occur, leading to a rise in deaths from early March 2020, behavioural adaptation by the general population might have reduced contact rates, resulting in a lower $R_0$. During 2020 Mogadishu was not reported to be affected by large-scale armed conflict, influx of displaced people or food insecurity, as in previous phases of the protracted crisis in Somalia. Differential under-ascertainment of burials over time in the satellite imagery analysis may therefore provide a more plausible explanation. Burials in the deceased's village of origin outside the capital and a potential decrease in the number of these burials (and thereby an increase of burials within city limits) due to mobility restrictions could have also played a role. It is also plausible that some of the excess mortality, especially its long plateau over the summer of 2020, was due to the NPIs themselves and other socio-economic disruptions due to the pandemic, though they cannot explain the early rise in excess burials predating NPIs.

Our best fits were obtained at IFR values higher than if age-specific IFRs were identical to consensus estimates from high-income countries. With the above qualifications in mind, this finding can be interpreted as supporting a higher IFR than expected from age demographics only, which could be due to untreated comorbidities and limited access to treatment. Raising the assumed population-average IFR value from 0.35% to 1.6% shifted the date of introduction estimates by approximately one month (Figure 2 and Figure 3). Even higher IFR values than 1.6 would require the IFR for those older than 75 to reach unrealistically high levels (over 70%). However, while in order to avoid over-parameterisation we shifted (the logit of) age-specific IFR values uniformly, it is possible that this shift was in reality differential and more pronounced for the larger younger age groups[30]. This could raise the population-average IFR above 1.6%, potentially shifting the estimated date of introduction to a later date.

The $R_0$ estimates between 1.4-1.7 are substantially lower than consensus estimates for SARS-CoV-2 for China[46,47], Europe[25,48] or the United States[49]. There is no empirical contact matrix or real-time mobility data available for Somalia, it is therefore possible that contact structures are in reality somewhat different from the projected contact matrix[29] for its neighbour (Ethiopia) that we used for model fitting, contributing to a lower reproduction number. Another possibility is that susceptibility to infection in younger individuals (*Extended data, SI Figure 5*[22]) is lower than estimates inferred for middle and high-income countries[24], i.e., that the lower median age in Somalia reduced the $R_0$ further. Relatively high ventilation of houses and proportion of time spent outdoors due to warm weather may also have reduced transmissibility. Other factors such as cross-immunity have also been proposed[50].

Finally, we note that our model fitting resulted in attack rates between 15–50% (*Extended data,* SI Figure 12[22], which would have left a large pool of susceptibles for a second wave to develop if the reproduction number increased due to introduction of more contagious new variants in late 2020. Indeed, a reportedly sharp pandemic wave was observed in Somalia[51] from late February to May 2021.

In summary, our analysis was based on the assumption that the rise in excess mortality observed via satellite imagery of cemeteries in Mogadishu starting early 2020 was due to deaths caused by SARS-CoV-2 infection. Under this assumption, model fitting of the time series of deaths suggests that SARS-CoV-2 could have been introduced to Somalia's capital substantially earlier than previously thought and had a basic reproduction number ($R_0$) lower than consensus estimates from middle- and high-income countries. To obtain date of introduction estimates no earlier than mid-November 2019 we had to assume population-wide IFR values of 0.7% or above, substantially above the value obtained by applying age-specific IFR estimates from high income countries to Somalia's age structure (0.15%). Even assuming thousands of external introductions of SARS-CoV-2 the median date of introduction estimate is shifted to mid-December only (*Extended data,* SI Figure 11[22]). Alternatively, there was another source of sustained excess mortality in Mogadishu from March to August 2020. If these excess deaths were indeed due to SARS-CoV-2 infections, this raises several questions about the pathogen's introduction and the true burden of the pandemic in Somalia and countries in the region. Further investigation of mortality trends and SARS-CoV-2 epidemiology in Somalia and other low-income countries is warranted to paint a more conclusive picture, and help to better predict future waves of the pandemic in these settings.

## Data availability

### Underlying data

Zenodo: Source data for "Date of introduction and epidemiologic patterns of SARS-CoV-2 in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020". https://doi.org/10.5281/zenodo.5534769[32].

This project contains the following underlying data:

- 2018-04-23-2021-04-28-Somalia.csv (ACLED data on political violence)

- df_compare_report_satell.csv (satellite data on number of burials compared to reported deaths)

The .rds files below are from the MCMC fitting, dates showing the time window of fitting, the value after "ifr_increm" the increment to the logit of the IFR curve, and the value after "seedsize" the number of external introductions:

- fits_death_2020-03-02_2020-08-02_ifr_increm0_seedsize100.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm0_seedsize20.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm0_seedsize200.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm0_seedsize50.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm1_seedsize100.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm1_seedsize20.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm1_seedsize200.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm1_seedsize50.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2.5_seedsize100.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2.5_seedsize20.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2.5_seedsize200.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2.5_seedsize50.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2_seedsize100.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2_seedsize20.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2_seedsize200.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm2_seedsize50.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm3_seedsize100.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm3_seedsize20.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm3_seedsize200.rds

- fits_death_2020-03-02_2020-08-02_ifr_increm3_seedsize50.rds

- IFR_by_age_imperial.csv (IFR estimates from [3] Imperial study)

- IFR_estimates_Sandmann2021.csv (IFR estimates from [26])

- out_bdr_daily_estimates.csv (Processed satellite data on number of burials)

- suscept_clinfract_posteriors_davies2010.csv (estimates on clinical fraction and susceptibility to infection from [24])

### Extended data

Zenodo: Extended data for "Date of introduction and epidemiologic patterns of SARS-CoV-2 in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020". https://doi.org/10.5281/zenodo.5525349[22].

This project contains the following extended data:

- SI.pdf (Supplementary material for "Date of introduction and epidemiologic patterns ofSARS-CoV-2 in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020")

- SI_Fig1.tiff (SI Figure 1. Burial rate per 10.000 person-days inferred from satellite imagery by interpolating between data points provided by satellite images (for details of methods see the accompanying paper[18]). The color shading represents estimates with low and high population denominators.)

- SI_Fig2.tiff (SI Figure 2. A. Burials by cemetery. Barakaat 1 cemetery was mostly filled up by January 2020 and replaced by its extension Barakaat 2. B. Dates for which satellite imagery was acquired.)

- SI_Fig3.tiff (SI Figure 3. '*StringencyIndex*' from the Oxford COVID-19 Government Response Tracker database. We divided the months of 2020 from March into four periods (separated by solid horizontal lines) and took the average of each. The fourth period is outside the time window of model fitting. The values are equal to the relative reduction in contacts (transmissibility) if the effectiveness of the NPIs (*NPI_scale*) was 1 (100%).)

- SI_Fig4.tiff (SI Figure 4. Weekly fatalities due to political violence in the Banadir region (from ACLED database) compared to the number of burials, going back to January 2019. The infrequency of satellite imagery analysed before 2020 does not enable identification of acute peaks in burials that would follow a mass casualty incident.)

- SI_Fig5.tiff (SI Figure 5. Clinical fraction and susceptibility by age group. Values from the literature[24] (green) were approximated by a piecewise linear function (red) made up of 3 sections (minimum, maximum and a line connecting them). The linearly approximated values were used as model parameters to minimise model complexity.)

- SI_Fig6.tiff (SI Figure 6. IFR estimates by age groups, using estimates from 26. Adjusted curves were calculated by taking the logit of the age-specific IFRs and adding the values 1, 2, 3.)

- SI_Fig7.tiff (SI Figure 7. Dynamic fits with all seed sizes and IFR estimates, fitting the period of 2 March to 24 August 2020. Labels show DIC values and median estimates for the date of introduction for the five different population-average IFR values that we tested.)

- SI_Fig8.tiff (SI Figure 8. Posterior distributions of the three fitting parameters (R0, introduction date,

NPI effectiveness (npi_scale)) generated by MCMC fitting, showing correlations between the parameters.)

- SI_Fig9.tiff (SI Figure 9. Model fitting restricted of the period 23 February to 13 April with four different seed sizes.)

- SI_Fig10.tiff (SI Figure 10. Model fitting of the period 23 February to 24 August 2020. Labels show DIC values and median estimates for the date of introduction by population-average IFR values.)

- SI_Fig11.tiff (SI Figure 11. Model fitting of the period from 02 March to 24 August with seed sizes up to 1000)

- SI_Fig12.tiff (SI Figure 12: Epidemic size for different fits. Cumulative attack rates for different seed sizes and IFR values, fitting the period 02/03/2020-24/08/2020.) Different IFR values lead to different estimates of $R_0$ and NPI_scale, resulting in different herd immunity thresholds and attack rates.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

## Analysis code
The code for model fitting and producing the figures is available at: https://github.com/mbkoltai/covid_lmic_model/tree/v1.0

Archived source code at time of publication: https://doi.org/10.5281/zenodo.5534763[19].

License: MIT

## Authorship information
The following authors were part of the Centre for Mathematical Modelling of Infectious Disease COVID-19 Working Group: Adam J Kucharski, Sebastian Funk, Rosalind M Eggo, Mark Jit, W John Edmunds, Amy Gimma, Stefan Flasche, Billy J Quilty, Samuel Clifford, Sam Abbott, James D Munday, Nikos I Bosse, Joel Hellewell, Hamish P Gibbs, Yang Liu, Nicholas G. Davies, Timothy W Russell, Christopher I Jarvis, Alicia Rosello, Carl A B Pearson, Kiesha Prem, Graham Medley, Gwenan M Knight, Akira Endo, Simon R Procter, C Julian Villabona-Arenas, Damien C Tully, Katherine E. Atkins, Sophie R Meakin, Rachel Lowe, Matthew Quaife, Oliver Brady, Kaja Abbas, Rosanna C Barnard, Frank G Sandmann, Kathleen O'Reilly, Mihaly Koltai, Yalda Jafari, William Waites, David Hodgson, Emilie Finch, Ciara V McCarthy, Rachael Pung, Paul Mee, Lloyd A C Chapman, Fiona Yueqian Sun, Stéphane Hué, Kerry LM Wong.

# References

1. Favas C, Jarrett P, Ratnayake R, *et al.*: **Country differences in transmissibility, age distribution and case-fatality of SARS-CoV-2: a global ecological analysis.** *medRxiv.* 2021.
   **Publisher Full Text**

2. **Report 34 - COVID-19 Infection Fatality Ratio Estimates from Seroprevalence.** Imperial College London. [cited 2021 Mar 31].
   **Reference Source**

3. Levin AT, Hanage WP, Owusu-Boaitey N, *et al.*: **Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications.** *Eur J Epidemiol.* 2020; **35**(12): 1123–38.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

4. Biccard BM, Joubert I, Kifle F, *et al.*: **Patient care and clinical outcomes for patients with COVID-19 infection admitted to African high-care or intensive care units (ACCCOS): a multicentre, prospective, observational cohort study.** *Lancet.* 2021; **397**(10288): 1885–94.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

5. van Zandvoort K, Jarvis CI, Pearson CAB, *et al.*: **Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study.** *BMC Med.* 2020; **18**(1): 324.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

6. Wiens KE, Mawien PN, Rumunu J, *et al.*: **Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Juba, South Sudan: a population-based study.** *medRxiv.* 2021; 2021.03.08.21253009.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

7. Gomaa MR, El Rifay AS, Shehata M, *et al.*: **Incidence, household transmission, and neutralizing antibody seroprevalence of Coronavirus Disease 2019 in Egypt: Results of a community-based cohort.** *PLoS Pathog.* 2021; **17**(3): e1009413.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

8. Mwananyanda L, Gill CJ, MacLeod W, *et al.*: **COVID-19 deaths detected in a systematic post-mortem surveillance study in Africa.** *medRxiv.* 2020; 20248327.
   **Publisher Full Text**

9. Van Elsland S, Watson O, Alhaffar M, *et al.*: **Report 31: Estimating the burden of COVID-19 in Damascus, Syria: an analysis of novel data sources to infer mortality under-ascertainment**. Imperial College London; 2020; [cited 2021 Mar 31].
   **Reference Source**

10. Karlinsky A, Kobak D: **Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset**. Davenport MP, Lipsitch M,Lipsitch M,Simonsen L, Mahmud A, editors. *eLife.* 2021; **10**: e69336.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

11. Watson OJ, Alhaffar M, Mehchy Z, *et al.*: **Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria.** *Nat Commun.* 2021; **12**(1): 2394.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

12. Adepoju P: **Nigeria responds to COVID-19; first case detected in sub-Saharan Africa.** *Nat Med.* 2020; **26**(4): 444–8.
    **PubMed Abstract** | **Publisher Full Text**

13. Nadeau SA, Vaughan TG, Scire J, *et al.*: **The origin and early spread of SARS-CoV-2 in Europe.** *Proc Natl Acad Sci U S A.* 2021; **118**(9): e2012008118.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

14. La Rosa G, Mancini P, Bonanno Ferraro G, *et al.*: **SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring.** *Sci Total Environ.* 2021; **750**: 141711.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

15. Koum Besson ES, Norris A, Bin Ghouth AS, *et al.*: **Excess mortality during the COVID-19 pandemic: a geospatial and statistical analysis in Aden governorate, Yemen.** *BMJ Glob Health.* 2021; **6**(3): e004564.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

16. Endris BS, Saje SM, Metaferia ZT, *et al.*: **Excess Mortality in the Face of COVID-19: Evidence from Addis Ababa Mortality Surveillance Program.** Rochester, NY: Social Science Research Network. 2021; Report No.: ID 3787447.
    **Publisher Full Text**

17. Elyazar IR, Surendra H, Ekawati LL, *et al.*: **Excess mortality during the first ten months of COVID-19 epidemic at Jakarta, Indonesia**. 2020; 2020.12.14.20248159.
    **Publisher Full Text**

18. Warsame A, Bashiir F, Freemantle T, *et al.*: **Excess mortality during the COVID-19 pandemic: a geospatial and statistical analysis in Mogadishu, Somalia.** *Int J Infect Dis.* 2021; **113**: 190–199.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

19. Koltai M: **mbkoltai/covid_lmic_model: Model fitting excess mortality data in Mogadishu v1.0 (v1.0).** *Zenodo.* 2021.
    **http://www.doi.org/10.5281/zenodo.5534763**

20. Division UNP: **wpp2019: World Population Prospects 2019.** 2020; [cited 2021 Apr 27].
    **Reference Source**

21. Qader SH, Lefebvre V, Tatem AJ, *et al.*: **Using gridded population and quadtree sampling units to support survey sample design in low-income**

settings. *Int J Health Geogr.* 2020; **19**(1): 10.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

22. Koltai M, Warsame A, Bashiir F, *et al.*: **Extended data for "Date of introduction and epidemiologic patterns of SARS-CoV-2 in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020"**. 2021.
    **http://www.doi.org/10.5281/zenodo.5527007**

23. Warsame A, Frison S, Gimma A, *et al.*: **Retrospective estimation of mortality in Somalia, 2014-2018: a statistical analysis**. 2020.
    **Reference Source**

24. Davies NG, Klepac P, Liu Y, *et al.*: **Age-dependent effects in the transmission and control of COVID-19 epidemics.** *Nat Med.* 2020; **26**(8): 1205–11.
    **PubMed Abstract** | **Publisher Full Text**

25. Davies NG, Kucharski AJ, Eggo RM, *et al.*: **Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study.** *Lancet Public Health.* 2020; **5**(7): e375–85.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

26. Sandmann FG, Davies NG, Vassall A, *et al.*: **The potential health and economic value of SARS-CoV-2 vaccination alongside physical distancing in the UK: a transmission model-based future scenario analysis and economic evaluation.** *Lancet Infect Dis.* 2021; **21**(7): 962–974.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

27. Cao B, Wang Y, Wen D, *et al.*: **A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19.** *N Engl J Med.* 2020; **382**(19): 1787–99.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

28. Linton NM, Kobayashi T, Yang Y, *et al.*: **Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data.** *J Clin Med.* 2020; **9**(2): 538.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

29. Prem K, van Zandvoort K, Klepac P, *et al.*: **Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era.** *PLoS Comput Biol.* 2021; **17**(7): e1009098.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

30. Van Gordon MM, Mwananyanda L, Gill CJ, *et al.*: **Regional comparisons of COVID reporting rates, burden, and mortality age-structure using auxiliary data sources**. *medRxiv.* 2021; 2021.08.18.21262248.
    **Publisher Full Text**

31. **COVID-19 Government Response Tracker**. [cited 2021 Mar 30].
    **Reference Source**

32. Koltai M: **Source data for "Date of introduction and epidemiologic patterns of SARS-CoV-2 in Mogadishu, Somalia: estimates from transmission modelling of satellite-based excess mortality data in 2020" [Data set]**. *Zenodo.* 2021.
    **http://www.doi.org/10.5281/zenodo.5534769**

33. Sanche S, Lin YT, Xu C, *et al.*: **High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2.** *Emerg Infect Dis.* CDC. 2020; **26**(7): 1470–1477.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

34. Jarvis CI, Van Zandvoort K, Gimma A, *et al.*: **Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK.** *BMC Med.* 2020; **18**(1): 124.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

35. Sapra I, Nayak BP: **The protracted exodus of migrants from Hyderabad in the time of COVID-19.** *J Soc Econ Dev.* 2021; 1–16.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

36. **ACLED 2020: The Year in Review**. 2021; [cited 2021 Apr 28].
    **Reference Source**

37. Menkir TF, Chin T, Hay JA, *et al.*: **Estimating internationally imported cases during the early COVID-19 pandemic.** *Nat Commun.* 2021; **12**(1): 311.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

38. Musa SS, Zhao S, Wang MH, *et al.*: **Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa.** *Infect Dis Poverty.* 2020; **9**(1): 96.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

39. Rice BL, Annapragada A, Baker RE, *et al.*: **Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa.** *Nat Med.* 2021; **27**(3): 447–53.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

40. Bloom JD, Chan YA, Baric RS, *et al.*: **Investigate the origins of COVID-19.** *Science.* 2021; **372**(6543): 694.
    **PubMed Abstract** | **Publisher Full Text**

41. Lochery E: **Somali Ventures in China: Trade and Mobility in a Transnational Economy.** *Afr Stud Rev.* 2020; **63**(1): 93–116.
    **Publisher Full Text**

42. Venkataraman M, Gofie SM: **The dynamics of China-Ethiopia trade relations: economic capacity, balance of trade & trade regimes.** *Bdg J Glob South.* 2015; **2**(1): 8.
    **Publisher Full Text**

43. Deslandes A, Berti V, Tandjaoui-Lambotte Y, *et al.*: **SARS-CoV-2 was already spreading in France in late December 2019.** *Int J Antimicrob Agents.* 2020; **55**(6): 106006.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

44. Basavaraju SV, Patton ME, Grimm K, *et al.*: **Serologic Testing of US Blood Donations to Identify Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)-Reactive Antibodies: December 2019-January 2020.** *Clin Infect Dis.* 2021; **72**(2): e1004–e1009.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

45. Kumar S, Tao Q, Weaver S, *et al.*: **An evolutionary portrait of the progenitor SARS-CoV-2 and its dominant offshoots in COVID-19 pandemic.** *Mol Biol Evol.* 2021; **38**(8): 3046–3059.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

46. Riou J, Althaus CL: **Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020.** *Euro Surveill.* 2020; **25**(4): 2000058.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

47. Kucharski AJ, Russell TW, Diamond C, *et al.*: **Early dynamics of transmission and control of COVID-19: a mathematical modelling study.** *Lancet Infect Dis.*
    2020; **20**(5): 553–8.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

48. Flaxman S, Mishra S, Gandy A, *et al.*: **Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.** *Nature.* 2020; **584**(7820): 257–61.
    **PubMed Abstract** | **Publisher Full Text**

49. Ke R, Romero-Severson E, Sanche S, *et al.*: **Estimating the reproductive number $R_0$ of SARS-CoV-2 in the United States and eight European countries and implications for vaccination.** *J Theor Biol.* 2021; **517**: 110621.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

50. Organization (WMO) WM, World Meteorological Organization (WMO): **First Report of the WMO COVID-19 Task Team : Review on Meteorological and Air Quality Factors Affecting the COVID-19 Pandemic (WMO-No. 1262)**. Geneva: WMO; WMO. 2021; 42.
    **Reference Source**

51. Salyer SJ, Maeda J, Sembuche S, *et al.*: **The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study.** *Lancet.* 2021; **397**(10281): 1265–75.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

# Open Peer Review

## Current Peer Review Status: ✓ ✓

---

**Version 2**

Reviewer Report 07 March 2022

https://doi.org/10.21956/wellcomeopenres.19622.r48813

✓ **Bhaswar Ghosh**

International Institute of Information Technology, Hyderabad, India

The authors present a modeling study regarding COVID-19 spread in Somalia. The study specifically uses death burial data to calculate the reproduction number $R_0$. Some of the concerns I expressed, are being addressed adequately.

1. They have used only death burial data to calculate the reproduction number since the daily infection data is limited.

2. They have provided an explanation of systematic change in $R_0$ with IFR.

3. They provided a detailed description of the figure 4 regrading the lines with different colors.

4. They have provided an explanation of possible mismatch between the simulation and data for figure 5.

I appreciate the authors for addressing the issues and providing explanation in the main text. I would recommend the article for an approval of indexing.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Computational Biology, Dynamical models, Biophysics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

**Version 1**

Reviewer Report 20 December 2021

https://doi.org/10.21956/wellcomeopenres.19063.r47064

? **Bhaswar Ghosh**
International Institute of Information Technology, Hyderabad, India

The article presents a method to estimate the infection rate and various other parameters based on the death data from satellite images of burial grounds in Somalia. The technique provides a quite important method to estimate the parameters from a different data source. Since the reported case numbers and deaths are heavily underestimated in many poor and middle-income countries, such alternative strategies would be highly pertinent for the current situation. I have a few concerns about the methods in the study.

1. The authors used only the death data to fit the epidemic model. This may underestimate the values of the $R_0$, since $R_0$ is the actual infection spreading rate. They may have used the reported case numbers also for simultaneously fitting both case numbers and deaths.

2. In Figure 2, a systematic reduction in $R_0$ value is observed with assumed fatality rate. They may provide some explanation for this either in the results section or in the discussion.

3. In Figure 4, it seems that the other solid lines (blue, cyan, green) are some estimates of the death rate from the data. This is not mentioned in the figure legends. The authors should mention the description of those lines in the legends to make it clear. In addition, there is a significant increase in death at 2020-8-03 which is not captured by the model, possibly due to the assumption of fixed time-independent parameter values.

4. In Figure 5, it is not clearly mentioned how the simulated curves change discontinuously at 2020-6-29, as it is mentioned that parameter values remain fixed at the mean estimated values.

5. In the methods section (Transmission model), the authors mentioned '*Both of these estimates are age-dependent, with the clinical fraction 29% in the 0-9y age group and 69% above 70 years, and susceptibility among individuals aged 0-19y half of that among adults*'. It is not clear what is the basis for this assumption or it is taken from any other study.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Computational Biology, Dynamical models, Biophysics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

> Author Response 15 Feb 2022
> **Mihaly Koltai**, London School of Hygiene & Tropical Medicine, London, UK
>
> Dear Dr. Ghosh,
>
> We thank you for your valuable comments concerning our manuscript on estimating the date of introduction of SARS-CoV-2 in Mogadishu (Somalia) based on satellite imagery of cemeteries. Please find our responses to the points raised below.
>
> 1) It is the case that we only fitted our transmission model to excess mortality data and not case numbers to derive the $R_0$ estimate. We would argue that this was the correct and indeed only way to proceed for two reasons.
>
> First, the transmission model we used (the COVIDM model) transforms infections to deaths using a delay distribution taken from the literature (a gamma distribution with a mean of 15 days). To the extent that there is a delay and a smoothing effect from infections to deaths this is captured by applying this delay distribution to the simulated infections.
>
> Second, using reported case numbers from Somalia was not a workable option in our case as reported cases only exceeded 10 cases on the 10th of April 2020, well after the rise in burials seen in our data, which started in late February. Therefore, fitting reported case numbers is inconsistent with our dataset and could not be used as a source of comparison, since the data of our interest, i.e.. excess mortality rising from late February is outside the earliest time point of reported cases.
>
> 2) We included an explanation on this in Results. Briefly, since we are fitting to deaths data only, the IFR is not a separately identifiable parameter and therefore we had to run separate fits with different assumed IFR values. When using the lowest assumed IFR (0.14%, see also SI Fig 6; the value was generated by applying age-specific IFR estimates from the literature to Somalia's age structure) this leads to higher $R_0$ estimates, as a larger number of infections need to be generated to be able to fit at all the number of deaths in the data. However we note that a) this fit at IFR=0.14% is of poor quality (see Figure 4) b) above this value of IFR the $R_0$ does not substantially change with the IFR, instead there is an inverse

relationship with the seed size, as a larger initial number of infections and a larger $R_0$ are exchangeable (within a certain range).

3) The solid lines in Fig 4 are generated by sampling from the posterior distributions obtained from MCMC fitting at different assumed values of the IFR. As noted in the previous point the IFR cannot be separately identified when relying on deaths data only, therefore different values had to be assumed and the MCMC fitting performed at these different input (IFR) values. The five solid lines correspond to the medians and the CI95 of the posterior distributions. We added this information to the figure legend.

The model does not fit well the increase in deaths from the 24th of August because we restricted the time window of model fitting to the period from 02/03/2020 to 24/08/2020. This is because our aim was the estimation of the date of introduction from the early rise of excess deaths. It is possible that the spike in deaths from the 24th of August could be fitted better if we increased the resolution of the 'StringencyIndex' from the Oxford COVID-19 Government Response Tracker as stringency started to fall in early August. We decided to rather introduce the stringency index with a lower resolution (four periods with fixed values) to minimise the number of exogenously introduced parameters.

4) The values for the stringency of interventions, which are introduced as scaling terms for the contact rates are fixed **within** the four periods that we aggregated the time series of 'StringencyIndex' into, however they change **between** these periods (SI Figure 3). Specifically the averaged stringency index falls from 0.6 to 0.26 on the 1st of July, causing the spike in infections (and later in deaths) for the three fits with the highest assumed IFR values.

5) The source for the age-specific curve of susceptibility and the clinical fraction is given in the previous sentence, it is reference [24] (Davies NG, Klepac P, Liu Y, et al.: Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med. 2020; 26(8): 1205–11). The age-dependent susceptibility and clinical fraction curves are shown in SI Fig 5.

We hope that we have answered all your comments and you will be able to approve the article. We are happy to answer any remaining concerns.

Sincerely yours,
Mihaly Koltai, on behalf of all authors

*Competing Interests:* No competing interests were disclosed.

Reviewer Report 19 October 2021

**Zindoga Mukandavire**

Emirates Aviation University, Dubai, United Arab Emirates

The authors used a dynamic transmission model to analyse a time series of excess deaths in Mogadishu (Somalia) using data extracted from satellite images of six main cemeteries. The study aimed at estimating the probable date of introduction of SARS-CoV-2, the basic reproduction number and the effect of non-pharmaceutical interventions. This is a very interesting study which uses novel approaches to understand COVID-19 transmission dynamics in countries with weak surveillance systems or crisis-affected settings. I have the following comments.

1. Consider shortening the title of the paper to read "*Date of introduction and transmissibility of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Mogadishu, Somalia: estimates from mathematical modelling*".

2. In the conclusion section of the abstract, it is not clear why estimates on transmissibility for Somalia are compared to European settings and not settings in Africa?

3. In data sources and statistical analysis section: Explain how you estimated that five smaller private and family-owned cemeteries accounted for < 20% of all burials. How did this affect the study findings?

4. How did the modelling results vary for seed sizes below 20?

5. The first column in Table 1, should be labelled "Parameter description".

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Applied Mathematics, Modelling Infectious Diseases

**I confirm that I have read this submission and believe that I have an appropriate level of**

**expertise to confirm that it is of an acceptable scientific standard.**

Author Response 20 Oct 2021

**Mihaly Koltai**, London School of Hygiene & Tropical Medicine, London, UK

Dear Dr. Mukandavire,

Thank you for your review of our paper and the constructive suggestions.
Responding to the points that you raised:

1) We will consider shortening the title in a second version.

2) Comparison was with European settings because more estimates were available for these. We will consider adding some more references on estimated $R_0$ in African settings.

3) This figure was based on the qualitative field work done by the colleagues at SDRI (Somali Disaster Resilience Institute) who verified the existence and functionality of burial sites. They concluded that these smaller private cemeteries accounted for a very small minority of burials. This conclusion was reached by considering the physical size of the cemeteries as well as by speaking to private individuals who were able to provide information on the burials within these sites (a limited number of prominent or very close family members were buried in these sites). When comparing this information with the information provided by the larger professionally managed cemeteries, our colleagues provided the <20% estimate. A more detailed description of data collection is now published in the adjoining study.

4) Seed sizes smaller than 20 would push back the estimate for the date of introduction by another 5-14 days, depending on the exact value of assumed seed size. We think that date of introduction estimates even earlier than the ones presented are less likely than seed sizes of 20 or above.

5) We can add the column name 'parameter description'.

*Competing Interests:* No competing interests were disclosed.